WOODHOUSE EXHIBIT 1

APPENDIX A

KADREY ET AL. V. META PLATFORMS, INC.

CASE NO. 23-CV-3417-VC

HEARING ON CROSS-MOTIONS FOR PARTIAL SUMMARY JUDGMENT

MAY 1, 2025

1

KEY ISSUES

These are summary judgment cross-motions. Each moving party must establish that there are no material disputed issues of fact and that they are entitled to judgment as a matter of law.

1. With respect to its fair use defense, Meta bears a "heavy burden" of proof.

2. Meta must establish fair use as to each copy made, and each use made of such copy.

3. Meta must establish that it appropriated only so much as was reasonably necessary to achieve its purpose.

4. Fair use is an equitable defense, and Meta must establish that its actions were consistent with its equitable obligations.

KEY ISSUE 1: WITH RESPECT TO ITS FAIR USE DEFENSE, META BEARS A "HEAVY BURDEN" OF PROOF

Meta **"*bears the heavy burden* of showing there are no genuine issues of material fact about whether its copying was fair."** *Nat'l Fire Prot. Ass'n v. UpCodes*, 753 F. Supp. 3d 933, 954 (C.D. Cal. 2024) (quoting *Ambat v. San Francisco*, 757 F.3d 1017, 1031 (9th Cir. 2014)) (emphasis added).

- "The Supreme Court and our circuit *have unequivocally placed the burden of proof on the proponent* of the affirmative defense of fair use." *Dr. Seuss Enters. v. ComicMix LLC*, 983 F.3d 443, 459 (9th Cir. 2020) (abrogated on other grounds) (emphasis added).

- "Regardless of how fair use is viewed," whether an affirmative defense or not, "it is clear that the burden of proving fair use is always on the putative infringer." *Lenz v. Universal Music Corp.*, 815 F.3d 1145, 1152–53 (9th Cir. 2016) (citing *Bateman v. Mnemonics, Inc.*, 79 F.3d 1532, 1542 n. 22 (11th Cir. 1996)).

- *Perfect 10*, 508 F. 3d. at 1158: "At trial, the defendant in an infringement action bears the burden of proving fair use."

- Congress "structured the [fair use] provisions as an affirmative defense." *Harper & Row*, 477 U.S. at 561; *accord Campbell*, 510 U.S. at 590.

KEY ISSUE 2: META MUST ESTABLISH FAIR USE AS TO EACH COPY MADE, AND EACH USE MADE OF SUCH COPY

- Each use of a copyrighted work must independently satisfy the requirements of fair use to avoid liability.

- As the Supreme Court and Ninth Circuit have repeatedly held, "the same copying may be fair when used for one purpose but not another" *Warhol Found. v. Goldsmith*, 598 U.S. 508, 533 (2023).

- *See Fioranelli v. CBS Broad. Inc.*, 551 F. Supp. 3d 199, 239 (S.D.N.Y. 2021) (finding some uses fair use and others not, where CBS used unauthorized footage of 9/11 across a variety of programming).

- The court held use for news programming was non-transformative and thus not fair use but held there was a genuine dispute of fact as to the transformativeness of those same works' use in political documentaries.

KEY ISSUE 3: META MUST ESTABLISH THAT IT APPROPRIATED ONLY SO MUCH AS WAS REASONABLY NECESSARY TO ACHIEVE ITS PURPOSE

Even if Meta were permitted to make some copies for some purposes, to carry its "fair use" burden of proof, Meta was required to establish (and, for purposes of its summary judgment motion, establish without material factual dispute) that each copy and each use was fair—including that Meta only copied and used what was "reasonably necessary" for its secondary use.

See Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 510–11 (2023).

KEY ISSUE 4: FAIR USE IS AN EQUITABLE DEFENSE, AND META MUST ESTABLISH THAT ITS ACTIONS WERE CONSISTENT WITH ITS EQUITABLE OBLIGATIONS.

- **"Fair use presupposes 'good faith' and 'fair dealing.'** The trial court found that The Nation **knowingly exploited a purloined manuscript**. Unlike the typical claim of fair use, The Nation cannot offer up even the fiction of consent as justification." *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 562 (1985) (cleaned up).

- "To invoke the fair use exception, an individual must possess an **authorized copy of a literary work**." *Atari Games Corp. v. Nintendo of Am. Inc.*, 975 F.2d 832, 844 (Fed. Cir. 1992).

- In evaluating "bad faith," the Ninth Circuit applies "the general rule that a party claiming fair use must act in a manner generally compatible with principles of good faith and fair dealing." *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1164 n.8 (9th Cir. 2007) (citing *Harper & Row*, 471 U.S. at 562–63).

PLAINTIFFS' MOTION FOR PARTIAL SUMMARY JUDGMENT

KEY UNDISPUTED FACTS SUPPORTING PLAINTIFFS' MOTION

1. Meta downloaded copies of Plaintiffs' copyrighted works from websites which, as Meta knew, offered illegal "pirated" copies of the works.

2. Meta concealed its use of pirated data and torrenting.

3. To facilitate its downloading of pirated copies of Plaintiffs' works, Meta also uploaded massive numbers of illegal copies of published works to pirated websites.

4. Meta considered and negotiated licenses to use copyrighted materials to train its AI models. Meta's competitors have entered into such licenses.

5. The reasons Meta downloaded pirated copies of copyrighted works included evaluating the value of the works and which works were worth licensing.

FACT 1: META DOWNLOADED FROM WEBSITES WHICH, AS META KNEW, OFFERED ILLEGAL "PIRATED" COPIES OF BOOKS

	Source*
Eleonora Presani: "I don't think we should use **pirated material**. I really need to draw a line there." Xavier Martinet: "We want to buy books and be the nice open people here . . . However to make it happen and not letting the bad guys win, we need to make a case – fast – and cut some corners here and there."	Ex. 18, at -730
Eleonora Presani: "I feel that using **pirated material** should be beyond our ethical threshold. . . . SciHub, ResearchGate, LibGen are basically like PirateBay or something like that, they are distributing content that is protected by copyright and they're infringing it."	Ex. 32, at -171
Nisha Deo: "**It's the piracy** (and us knowing and being accomplices) that's the issue"	Ex. 37, at -899
"Z-lib: initially a mirror of LibGen, but then evolved to a separate project. Now claims to have 23M books and 285B articles. **Banned multiple times**, but seems to be working currently. Worth investigating."	Ex. 56, at -792
Zuckerberg LibGen Approval Memo: "If there is media coverage suggesting we have used a **dataset we know to be pirated**, such as LibGen, this may undermine our negotiating position with regulators on these issues."	Ex. 61, at -246
Nikolay Bashlykov: "Not sure we can use Meta's IPs to load through torrents **pirate content**, ahah . . . Yeah, I think torrenting from a corporate laptop doesn't feel right"	Ex. 68, at -224
Todor Mihaylov asks: "Is libgen this" following a screenshot of a Google search result that read "**Libgen is not legal**, and using Libgen may open you up to legal ramifications." Melanie Kambadur replies: "haha yes, we suspect that some of our competitors are using it and we want to understand better performance/legal tradeoffs in using it as well as alternatives we could license"	Ex. 63, at -325-26
Jort Gemmeke: "Specifically, I need the list of libgen *titles* that match the words: 'stolen' 'pirated' or 'unauthorized.'"	Ex. 90, at -740-41

*Unless otherwise indicated, all Exhibit citations are to the declarations of Maxwell V. Pritt.

9

FACT 2: META TRIED TO CONCEAL ITS PIRATING AND TORRENTING

	Source
Todor Mihaylov suggested edit: Replace: "pirate" with "external"	Ex. 54, at -583
Zuckerberg LibGen Approval Memo states, "in no case would we disclose publicly that we had trained on libgen"	Ex. 61, at -245
Melanie Kambadur: "pushshift and libgen probably approved for us to use . . . Supposed to be on a 'need to know' basis."	Ex. 62, at -848
Nikolay Bashlykov: "Not sure we can use Meta's IPs to load through torrents pirate content, ahah . . . Yeah, I think torrenting from a corporate laptop doesn't feel right"	Ex. 68, at 224
David Esiobu: "Can you clarify why we can't use FB infra for this again?" Frank Zhang: "Avoiding risk of tracing back the seeder . . . From FB server" "there's an explicit decision to not involve [Data Infrastructure team]. We want to get AA ingesting done in stealth mode within our group"	Ex. 73, at -239
"We also used this dataset [Books3] for Llama 2 which was not shared externally."	Ex. 148, at 475
Melanie Kambadur sends an attachment showing Llama output admitting the model was trained on LibGen and asks a colleague "Any chance we can fix this behavior? Looks pretty bad even if it's hallucinating."	Ex. 149, at -518-24
Meta employees discussed concern over whether the public could "deduce" whether Books3 "was used in llama 2," Meta's first admittedly commercial model, but Meta had publicly disclosed use of Books3 to train Llama 1 as a purportedly "research-only" model.	Ex. 156, at -906

FACT 3: TO FACILITATE ITS DOWNLOADING, META UPLOADED MASSIVE NUMBERS OF ILLEGAL COPIES TO PIRATED WEBSITES

	Source
Meta Expert Barbara Frederiksen-Cross: "For the April-July 2024 time period . . . After deduplication, this data shows 134.6 TB of data downloaded from the Internet, and **40.42 TB** of data uploaded to the Internet." (emphasis added)	Ex. 163, n. 56 (Ex. 163, Frederiksen-Cross Second Rebuttal Report at 29 n.56 ("this data shows 134.6 TB of data downloaded from the Internet, and 40.42 TB of data uploaded to the Internet").
This downloading included over **650 copies** of Plaintiffs' works at issue in this case.	Ex. 72, Frederiksen-Cross Rebuttal Report, at 47 (Table 2 – Number of Downloaded Plaintiffs' Works Across Torrented Datasets).

11

FACT 4: META NEGOTIATED LICENSES FOR DATA, AND COMPETITORS HAVE ENTERED INTO LICENSES

	Source
Mar. 2023: "We should try to get all the big guys: Harper Collins, Simon & Schuster, Macmillan, Hachette, Penguin Random House."**	Ex. 41, at -554
Mar. 2023: Emails with ███████████████ about licensing opportunity.	Ex. 42
Mar. 2023: Strategy document discussing potential licensing deals for fiction with ██████████████████████████	Ex. 44, at -954
Mar. 2023: Meta contemplated book licensing spend would require $17 million, then $27 million.	Ex. 48, at -195
Mar. 2023: Book licensing spend contemplated as "reasonable" at $50 million.	Ex. 49, at -332
Mar.-Apr. 2023: Emails with ████████ about licensing opportunity.	Ex. 43
Apr. 2023: Book licensing spend contemplated at up to $200 million, because "It would be a travesty if our models suck because we didn't spend $100M on data."	Ex. 92, at -029-30
June 2024: "We estimate that around ~10-20% of total text data corpus from Llama-4 should come from licensing."	Ex. 97, at -192

**Red boxes indicate material currently filed under seal.

12

FACT 4: META NEGOTIATED LICENSES FOR DATA, AND COMPETITORS HAVE ENTERED INTO LICENSES

AI companies have executed dozens of licensing agreements for textual training data.

In September 2024, ██████████ **licensed for training data its use of ████ books for █████ per book for ██████ from** ████████████████

Ex. 139 at -761.

AI Licensing Deals for Textual Works
(As Publicly Announced and/or Reported by AAP Members)[55]

Licensee	Content Owner		Licensee	Content Owner
Amazon	The Associated Press		OpenAI	Hearst
Amazon	Business Insider		OpenAI	Le Monde
Amazon	Condé Nast		OpenAI	News Corp
Amazon	Forbes		OpenAI	Prisa Media
Amazon	Hearst		OpenAI	Schibsted Media Group
Amazon	Politico		OpenAI	TIME
Amazon	Reuters		OpenAI	Vox Media
Amazon	Time		Perplexity	Adweek
Amazon	USA Today		Perplexity	The Independent
Amazon	The Washington Post		Perplexity	The Los Angeles Times
Amazon	Vox		Potato	Wiley
Dow Jones	The Associated Press		ProRata.ai	Adweek
Dow Jones	The Wall Street Journal		ProRata.ai	The Atlantic
Dow Jones	The Washington Post		ProRata.ai	Atlas Obscura
LexisNexis	The Associated Press		ProRata.ai	Arena Group
Meta	Reuters		ProRata.ai	Axel Springer
Microsoft	Axel Springer		ProRata.ai	Buzzfeed
Microsoft	Financial Times		ProRata.ai	DMG Media Group
Microsoft	HarperCollins		ProRata.ai	Financial Times
Microsoft	Hearst		ProRata.ai	Fortune
Microsoft	Reuters		ProRata.ai	Guardian Media Group
Microsoft	USA Today		ProRata.ai	Hello!
Mistrial	Agence-France-Presse		ProRata.ai	Mediahuis
OpenAI	American Journalism Project		ProRata.ai	Mumsnet
OpenAI	The Associated Press		ProRata.ai	News/Media Alliance (on behalf of its members)
OpenAI	The Atlantic		ProRata.ai	Prospect
OpenAI	Axel Springer		ProRata.ai	Reach PLC
OpenAI	Axios		ProRata.ai	Sky Media Group
OpenAI	Condé Nast		Confidential	Taylor & Francis
OpenAI	Dotdash Meredith		Confidential	Wiley
OpenAI	Financial Times		Confidential	Wiley
OpenAI	GEDI		Confidential	Wiley
OpenAI	Guardian Media Group		Confidential	Wiley

Ass'n of Am. Publishers Amicus Brief, Dkt. 535 at 12

**Red boxes indicate material currently filed under seal.

13

FACT 5: META USED PIRATED COPIES OF BOOKS TO EVALUATE WHICH BOOKS HAD VALUE AND WHICH BOOKS IT NEEDED TO LICENSE

	Source
Oct. 2022: "For now we have agreed to use [LibGen] just for pure exploration to see if there is value. If we find value, then we will set up proper licensing agreement."	Ex. 32, at -171
Dec. 2022: "The team was initially evaluating the usage of copyrighted work such as libgen fiction and non-OSS github. Results obtained in December show that the upside from using these is negligible, while the legal limitations in order to train a LLM for widespread internal use are significant. The team decided not to include these datasets in the final training."	Ex. 35, at -092
Apr. 2023: Meta generated a spreadsheet that cross-referenced the millions of works it torrented from LibGen against publishers' known book catalogs.	Ex. 75
Apr. 2023: "I don't think we need to proceed" with licensing negotiations "with ███████ at this point: ████████ overlaps with up to 90% of content in LibGen" "Do we still want to buy ██████████ despite the similarity?"	Ex. 56, at -018
June 2024: "Cross reference books we have collected from Anna's Archive and their publishers. Identify publishers from whom we have collected minimal or no book data." "We are still in the process of performing a query in union with what we collected from Anna's Archive & spidermate and **will only** go after the publishing companies with the largest delta of missing content" (emphasis in original).	Ex. 97, at -197-98

14

THERE ARE NO FAIR USES FOR PIRACY

<u>No</u> case has ever approved downloading illegal copies through file sharing networks as fair use.

- A fair-use defense to "downloading or uploading of [a] copyrighted work" via a P2P network is "baseless," and does not present "a close and difficult case." *Glacier Films (USA) v. Turchin*, 896 F.3d 1033, 1043 (9th Cir. 2018).

- The fair-use factors *"weigh against application of the fair use doctrine to cases involving Internet piracy . . . .* **It is preposterous to think that Internet piracy is authorized by virtue of the fair use doctrine."** *United States v. Slater*, 348 F.3d 666, 669 (7th Cir. 2003).

- "In some cases, no analysis is required; it is obvious, for example, that downloading and distributing copyrighted music via peer-to-peer systems does not constitute fair use." *In re DMCA § 512(H) Subpoena to Twitter*, 608 F. Supp. 3d 868, 879 (N.D. Cal. 2022) (Chhabria, J.).
  - Courts have devoted time and resources towards litigating criminal and civil cases involving pirated websites, including the shadow libraries at issue in this case.
  - The DMCA was adopted, in part, to combat the harms and proliferation of digital piracy; That piracy is incompatible with the copyright laws is beyond dispute.

15

META CITES NO CASE WHERE ILLEGALLY OR WRONGFULLY OBTAINING A COPYRIGHTED WORK WAS FAIR USE

None of the cases Meta cites for its assertion that its fair use defense survives even though its initial copy was made from an illegal, "pirated" version support Meta's claim.

- *None* of *Campbell, Google v. Oracle,* or *Los Angeles News Service* (Meta Reply at 1-2, 10) **even involve or address the issue.**

- In *Sony Comput. Ent. Inc. v. Connectix,* 203 F.3d at 608 (Meta Reply at 2), the page cited by Meta **does not address the issue.** The decision noted that "early in the development process" an engineer had "downloaded from the internet" a copy of Sony's software *Id.* at 601. However, **the downloaded copy "was not used** to develop" the emulator software at issue because the defendant abandoned it when it realized the copy was a Japanese language version. *Id.* **The court noted that the defendant "purchased" the copy it actually used.** *Id.*

  - Because of that abandonment and because *Connectix* was a case for injunctive relief, **the Ninth Circuit did not address the downloaded copy in its opinion.** *Id.* at 599. ("Any other intermediate copies made by Connectix do not support injunctive relief, even if those copies were infringing.").

  - In fact, in its own brief opposing Sony's Petition for Certiorari, Connectix wrote, **"the Ninth Circuit did not decide whether Connectix's use of the downloaded copy was an infringement,"** but rather concluded that even if it was, it did not justify injunctive relief. Respondent Connectix's Brief in Opposition, *Sony v. Connectix,* 2000 WL 33999755, at *3 n.2 (Aug. 2, 2000), cert. denied, 531 U.S. 871.

16

META CITES NO CASE WHERE ILLEGALLY OR WRONGFULLY OBTAINING A COPYRIGHTED WORK WAS FAIR USE

- *Perfect 10* involved Internet image indexing that "only incidentally indexes infringing websites." 508 F.3d at 1164 n.8.

- In *Kelly*, the search excluded websites once it was informed they contained unauthorized copyrighted material. *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 816 (9th Cir. 2003).

- These cases are far removed from the targeted piracy Meta engaged in here.

17

THERE ARE NO FAIR USES FOR PIRACY

Meta agrees, as it told regulators in 2023, after this case was filed:

"[O]ne of the most critical elements of the balance between rightsholder interests and innovation is the **doctrine of fair use**. Courts have applied that doctrine to lay the legal groundwork for revolutionary new technologies like internet search, **while declining to sanction more exploitative technologies, like unauthorized file sharing** and unlicensed media clip services."

Comments of Meta Platforms, U.S. Copyright Office, Dkt. No. 2023-06 at 11
(filed Oct. 30, 2023) (emphasis added).

18

PIRACY & PARODY: CONSIDER THE COMEDIAN

Under Meta's theory of "fair use," a comedian could pirate the complete Harry Potter series to develop a political parody of J.K. Rowling and, if that later parody was deemed "fair use," the comedian's earlier piracy would be absolved.

- Indeed, under Meta's theory a comedian could not only pirate the complete Harry Potter series from an illegal website but *also upload pirated data to fellow peers* in the swarm in the process. Because her later parody was deemed "fair use," all earlier infringing conduct would be absolved.

- Meta's interpretation of "fair use" has no precedent: to apply the doctrine so broadly would contravene decades of case law, sound policy, and common sense.

19

META'S BOB DYLAN DEFENSE

Steal a little and they throw you in jail

Steal a lot and they make you king

- Bob Dylan, "Sweetheart Like You"

BMG Music v. Gonzalez affirmed the district court's grant of summary judgment to record companies for copyright infringement and its award of statutory damages against **a single P2P user for downloading 30 songs**. 430 F.3d 888 (7th Cir. 2005).

Meta downloaded an "**astonishingly large volume**" of copyrighted books, including via P2P networks, while contending "it does not matter whether Meta downloaded datasets containing 'pirated' books from a third-party who lacked authorization to distribute them . . . " Meta Mot. at 31, 18.

20

META KNEW THAT THE SOURCES FROM WHICH IT DOWNLOADED PLAINTIFFS' WORKS WERE ILLEGAL, BANNED, CRIMINAL ENTERPRISES

Multiple Courts have issued permanent injunctions against Library Genesis.

- *Cengage Learning, Inc. v. Does 1-50 d/b/a Library Genesis*, Case No. 23-cv-8136-CM, Dkt. 36 (S.D.N.Y. Sept. 24, 2024):

    - "Defendants have **willfully infringed** Plaintiffs' copyrights . . . Plaintiffs have been irreparably harmed as a result of Defendants' unlawful conduct" *Id.* (emphasis added).

- *Elsevier Inc. v. www.Sci-Hub.org*, 2015 WL 6657363, at *4 (S.D.N.Y. Oct. 30, 2015):

    - "The Defendants cannot be legally harmed by the fact that they cannot **continue to steal the Plaintiff's content**, even if they tried to do so for public-spirited reasons." (cleaned up) (emphasis added).

DOJ is criminally prosecuting administrators of Z-Library – a shadow library Meta knew was banned, but from which it nonetheless torrented 49 TB of data.

- https://www.justice.gov/usao-edny/pr/two-russian-nationals-charged-running-massive-e-book-piracy-website

The DMCA was adopted in part to combat the kind of piracy Meta engaged in here.

- S. Rep. No. 105-190, at 8 (1998); H.R. Rep. No. 106-216 (1999).

21

META TORRENTED ALL OF PLAINTIFFS' WORKS AND UPLOADED MORE THAN 40 TB OF DATA TO THIRD PARTIES

Meta's expert also found Meta uploaded—i.e. it *reproduced and distributed to third parties*—over **40 TB** of the 134.6 TB of pirated material it torrented in April - June 2024.

- There is no dispute that Meta used a peer-to-peer file sharing network in 2024 to:

  - Copy all of Plaintiffs' works, and over 650 copies of those works, without authorization. Ex. 72, Frederiksen-Cross Rebuttal Report, at 47 (Table 2 – Number of Downloaded Plaintiffs' Works Across Torrented Datasets)

  - Distribute over 40 TB of the 134.6 TB of pirated material. Dkt. 492 at ¶ 29; Ex. 163 at 29.

- Meta proffers *no factual evidence* that none of Plaintiffs' copyrighted material was distributed.

Meta *also* torrented the entirety of the LibGen Fiction library in Fall 2022.

- This dataset contained 204 copies of Plaintiffs' works. Butterick Decl. ¶ 6.

- At this time, Meta did *not implement any measures to prevent distribution* in either the leeching or seeding stages; Meta produced logs showing *uploading* of this set, but not which works were or were not uploaded.

22

META CHOSE TO DISTRIBUTE PIRATED DATA TO SPEED UP ITS ACQUISITION OF THAT DATA

- Meta's expert admitted that Meta could have completely prevented the distribution of pirated data during the leeching and seeding phases of torrenting through simple changes to the libtorrent client.

  - Ex. 67, 107:19-108:18 (admitting the default torrent configuration allows for simultaneous uploading and downloading); 246:13-247:12 & 248:11-23 (noting there are ways to configure the torrent to eliminate uploading, but Meta did not do so). *See also* Ex. 100, 142:6-13 (Choffnes Trans., 142:6-13 (discussing whether it is possible to download via torrent without uploading while you do)).

- Instead, **Meta opted to distribute pirated data near-continuously for almost three months**. Frederiksen-Cross Trans., 41:1-22 (discussing Meta's use of the default settings for libtorrent) & 44:8-45:15 (same).

- Meta's expert confirmed that "peers, in order to optimize the effectiveness and the speed of transfer, will preferentially upload to peers who are downloading to them." Ex. 67, Frederiksen-Cross Trans., 130:16-19 ("Q. What do you mean by "tit for tat way that BitTorrent works? A. . . . And also that peers, in order to optimize the effectiveness and the speed of transfer, will preferentially upload to peers who are downloading to them.").

- Therefore, **Meta achieved faster downloading speeds by distributing pirated data while torrenting it**.

23

SCOPE OF META'S DISTRIBUTION OF PIRATED MATERIAL

- Plaintiffs' expert Dr. David Choffnes has "never seen an individual instance of an entity downloading or uploading this much data." Ex. 100, 140:9-16.

- "Between April 5th and April 11th alone, Meta uploaded 28 terabytes of data and downloaded 170 terabytes of data using those [Amazon AWS] EC2 instances." Ex. 100, 138:6-10.

24

META CONTRIBUTED MATERIAL SUPPORT TO PIRACY

- **Meta's expert testified that torrenting peers contribute "bandwidth, content, storage, and processing power of the entire network."** Ex. 67, 101:7-12 (quoting paragraph 44 of Frederiksen-Cross Rebuttal Report).

- **Meta's contributions were enormous.**

  - Meta used six virtual computers that were running near continuously for multiple months to download and upload data from Anna's Archive in 2024. Ex. 34, 52:7-25 (Clark 30(b)(6) (Torrenting), 52:7-25 (discussing Meta's torrenting of data from Anna's Archive).

  - Meta's torrenting bandwidth was "18 gigabits per second, which is a very fast link." Ex. 100, Choffnes Trans., 82:25-83:4 (discussing the factors that influence download time for a torrent).

25

TORRENTING EVIDENCE

Time	Source(s) & Content Type	Method / Location	Facts in Record re: Volume	Produced Log/Data Evidence	Missing Evidence	Testimony re: Unproduced Evidence
2022	LibGen (Fiction, SciTech – nonfiction, scientific articles)	Torrent to Meta H2 cluster (Penguin Computing, Utah)	~2.2 million files; 204 Copies of Plaintiffs' Works in Download	Upload and Download activity logged.	No data usage or command history. Original download content missing	Ex. 98 (55:23-56:5) (noting that the first torrented copy of LibGen has since gone missing).
2023	LibGen (SciMag – NonFiction, Scientific Articles, Fiction)	Torrent to Meta "Devserver"	~80 TB Downloaded	N/A	Evidence of download logs being created in code, but not produced. No data usage or command history.	Ex. 98 (117:1-120:20) "no logs had been located" because "it was outside the [one-year] retention period," and "it was "possible to prevent such overwriting from taking place."
2024	Anna's Archive (Internet Archive, Z-library); LibGen (Sci-Tech – Nonfiction, Scientific Articles); Duxiu.	6 AWS EC2 Instances	134.6 TB Download; 40.2TB Upload; 666 Copies of Plaintiffs works in Download (excluding Duxiu)	Data usage produced from AWS/ Amazon instances	Evidence of download logs being created in code, but not produced. No command history.	Ex. 98 (196:7-200:15 (testifying that "instances" (i.e. virtual computers) were also "destroyed" sometime between "October and Early November" (2024).

- Meta cites the ESI Order to argue it was not required to preserve logs. Reply at 5 (citing Dkt. 100-1, at 7).
- However, Meta fails to cite what follows: "this category will not include sources known to contain data that was used to train any large language model related to issues raised in this litigation or any dataset or records describing that data (hereinafter collectively 'LLM Data')." Dkt. 100-1, at 8.

26

LATE PRODUCTION TIMELINE

Event / Deadline	Date of Production	Production Type	Size	Notes
Close of Fact Discovery – December 13, 2024	Meta_Kadrey_029 Produced on December 13, 2024		2,404 Documents	On last night of fact discovery ending, Meta produced some of its hottest documents, including involving Mark Zuckerberg
MZ Depo – December 17, 2024				
	Meta_Kadrey_034–037 Produced Between February 10–close of 12, 2025	Sequestered Documents	1,348 Total Documents	Meta produced many more of its hottest documents long after the close of discovery
Close of Expert Discovery – February 26, 2025				
Response to MSJ – March 24, 2025				
	Meta_Kadrey_039–044 Produced Between March 19 – April 1, 2025	Torrenting	282 Total Documents	Meta produced 282 documents in response to RFP on torrenting searching across only 4 custodians.
King Deposition – April 2, 2025	April 2, 2025	Deposition Exhibit		Meta produced for the first time a summary document of key admissions.
	Reproduction of Meta_Kadrey_042 Produced on April 29, 2025	Reproduction	79,212 Documents	Meta reproduced a document that due to its error did not open properly revealing full logs of 2022 LibGen Fiction downloads and uploads.
MSJ Hearing – May 1, 2025				

27

THE NINTH CIRCUIT ADHERES TO A "DEEMED DISTRIBUTED" RULE FOR MATERIAL OFFERED ON P2P NETWORKS

Evidence that a defendant made a copyrighted work available to the public on a P2P network is sufficient to show distribution

- *See Perfect 10*, 508 F.3d at 1162, directly endorsing Napster's holding that participation in a P2P network like Napster, which makes the peer's "collections available to all other Napster users" is consistent with a "deemed distribution" rule. See *Perfect 10*, 508 F.3d at 1162 (citing *Hotaling* and *Napster*).

- *Napster* held that "users who upload *file names*" to Napster's "search index," as opposed to actual transfer of files, "violate plaintiffs' distribution rights," while "Napster users who *download* files containing copyrighted music violate plaintiffs' reproduction rights." 239 F.3d at 1014.

- *See also Hotaling v. Church of Jesus Christ of LDS*, 118 F.3d 199, 203 (4th Cir. 1997): An "offer to distribute [a] work" by "mak[ing] the work available" to the public is "considered distribution within the meaning of § 106(3)"; under a more restrictive rule requiring proof of actual distribution, "a copyright holder would be prejudiced by a library that does not keep records of public use, and the library would unjustly profit by its own omission."

28

THE "DEEMED DISTRIBUTED" RULE IN *PERFECT 10* AND *NAPSTER* IS CONSISTENT WITH INTERNATIONAL IP TREATIES AND THE "MAJORITY OF CASES"

SA Music v. Amazon observed that the **U.S. Copyright Office, treaties to which the U.S. is a signatory, and "the majority of cases analyzing the issue" provide that "the distribution requirement of § 106(3) is satisfied when a copyrighted work is made available for downloading through a file-sharing network** – as one might do on a peer-to-peer network such as Napster." 2020 WL 3128534, at *2–3 (W.D. Wash. June 12, 2020) (emphasis added).

> "In general, **where a party offers members of the public access to a work in the form of a download, the offer implicates the right of distribution . . .** the statutory language, context, and legislative history all indicate that **Congress intended to reserve to copyright owners the right to determine whether and how their works are made available to the public in copies, including digital files...."**

Copyright Office, "The Making Available Right in the U.S." (Feb. 2016) (quoted in *SA Music*).

29

META'S AUTHORITY ON DEEMED DISTRIBUTION IS INAPPOSITE AND OUTDATED

VHT, Inc. v. Zillow Group, Inc., 918 F.3d 723, 736 (9th Cir. 2019): Was not a P2P case, and held only that "making available" is not sufficient for violation of *display* rights—not distribution rights—when underlying proof shows there was no display, when the jury was never instructed on the "made available" theory, and plaintiff did not raise the issue in proposed jury instructions or in objections to the final jury instructions.

EVOX Prods., LLC v. Verizon Media Inc., 2021 WL 3260609, at *3–4 (C.D. Cal. May 5, 2021): Was not a P2P case, failed to consider the discussion in *Perfect 10* affirming the *Napster* rule for P2P "making available" violations, and was decided at the pleading stage on the ground that the plaintiff did not allege any non-conclusory facts concerning distribution.

Atlantic Recording Corp. v. Howell, 554 F. Supp. 2d 976, 983 (D. Ariz. 2008):

- Is inconsistent with *Perfect 10*'s affirming the *Napster* rule for P2P "deemed distributed" violations.
- Was based on a purported contradiction between *Perfect 10* and *Napster* on the "deemed distributed" rule even though *Perfect 10* specifically cited *Napster* with approval for that rule in the P2P context, in parallel with *Hotaling*.
- Is inconsistent with *SA Music*, decided twelve years after *Howell*, which concluded that the U.S. Copyright Office, international treaties to which the U.S. is a signatory, and "majority of cases analyzing the issue" applied the "deemed distributed" rule.

30

META'S MASS PIRACY CANNOT BE FAIR USE AS A MATTER OF LAW

- **To deny Plaintiffs' affirmative motion would condone conduct no court in history has ever found to be fair use.**

  - It would obliterate the right of creators to control access to their works and by permitting commercial actors to bypass the market for access thus undermining the very economic incentives copyright was designed to protect

  - It would mean that a trillion-dollar company can engage in voracious piracy so long as some of the later use of the pirated data is "transformative."

  - **"[C]opiers . . . cannot ask courts (and juries) to second-guess the market and call wholesale copying 'fair use' if they think that authors err in understanding their own economic interests or that Congress erred in granting authors the rights in the copyright statute."** *BMG Music v. Gonzalez*, 430 F.3d 888, 891 (7th Cir. 2005). (Easterbrook, J.)

- Further, it would endorse a position that centralized illegal libraries, which exist only through participation of peers willing to upload pirated data, have a legitimate purpose even though federal agencies, courts, and civil litigants are simultaneously pouring resources into trying to shut them down.

31

META'S MOTION FOR PARTIAL SUMMARY JUDGMENT

FAIR USE IS A FACT SPECIFIC INQUIRY RARELY SUITABLE FOR SUMMARY JUDGMENT

- Fair use is a "fact specific inquiry for which summary judgment is ill-suited," particularly where the disputed use lacks any directly applicable precedent, like here. *See Ass'n of Am. Medical Colleges v. Cuomo*, 928 F.2d 519, 524 (2d Cir. 1991).

- *See In re DMCA*, 608 F. Supp. 3d 868, 879 (N.D. Cal. 2022) (Chhabria, J.) (noting that apart from "obvious" cases like "downloading and distributing copyrighted music via peer-to-peer systems," "[t]he fair use analysis is fact intensive").

UNDISPUTED FACTS WARRANT DENIAL OF META'S SUMMARY JUDGMENT MOTION

- Meta's torrenting activity led to dozens of terabytes' worth of copyrighted works being shared.

- Meta's 2022 LibGen dataset was used to determine data licensing strategies. However, due to legal concerns, Meta never used this dataset to train any Llama model. Ex. 32, at -171.

- Meta's 2023 LibGen dataset was used to cross-reference LibGen titles with publishers' catalogues to determine data licensing strategies.

- Once Meta realized that its 2023 LibGen dataset greatly overlapped with ███████ catalogue, Meta called off licensing with ███████ for AI training data.

- All Plaintiffs authored literary works.

- Meta's Llama models are commercial.

- Meta made full copies of Plaintiffs' works.

- Meta employees actively concealed their use of pirated training data from the public out of concern that it would undermine their credibility with regulators.

- Meta destroyed or overwrote computers, logs, and other records that evidenced its torrenting activities in 2023 and 2024.

DISPUTED FACTS CONCERNING EVERY USE OF EVERY COPY OF PLAINTIFFS' BOOKS PRECLUDE GRANTING META'S SUMMARY JUDGMENT

The following factual disputes require denial of Meta's summary judgment with respect to _each_ of the copies that Meta made of Plaintiffs' works:

- Whether Meta copies the expressive elements of Plaintiffs' works _and_ for what purpose.

- Whether each copy, and each use of each copy, that Meta made of Plaintiffs' works was "reasonably necessary" to accomplish Meta's "new purpose."

- Whether "software products like Llama , . . . can be developed only through use of intermediate digital copies" of Plaintiffs' works. Meta Reply Brief at 18, n. 14.

DISPUTED FACTS CONCERNING MARKET IMPACT ALSO PRECLUDE GRANTING META'S SUMMARY JUDGMENT

- Whether Plaintiffs were denied a sale as the result of Meta's unauthorized copying.

- Whether the only market at issue is the market for licensing books for use as AI training data.

- Whether there is a "traditional" market that was affected by Meta's copying.

- Whether there is a "reasonable" market that was affected by Meta's copying.

- Whether there is a "likely to be developed" market that was affected by Meta's copying.

- Whether, if there is a market failure, Meta's use of pirated copies materially contributes to that failure.

- Whether Meta reproduced any of Plaintiffs' works by uploading them to P2P sites.

- Whether and to what extent Meta's copying of Plaintiffs' works contributes sufficient benefit to the public to outweigh the harms.

META MADE MANY DIFFERENT COPIES OF PLAINTIFFS' WORKS WITH DIFFERENT USES

Description	Non-Exhaustive Examples of Meta's Uses
Books3 from the Pile (2022)	Downloaded and stored; some were used for LLM training; others for research, evaluation, or benchmarking. Meta admits to downloading and storing full copies; not all were used for training. Exs. 37-39.
2022 Download of LibGen Fiction	Initial 2022 LibGen copy was used for "exploration" and not for LLM training; later copies were used for other purposes, like developing an internal only model, and assessing the usefulness of LibGen by other teams at Meta. Exs. 31, 35-36.
2024 Download of Anna's Archives (Z-Lib, Internet Archive, and LibGen)	Training data and "gap" analysis. Anna's Archive aggregates LibGen, Z-Lib, IA. See Exs. 17, 64-66.
2023 Download of LibGen (Science/Non-Fiction)	Used to evaluate whether Meta "still need to buy ▮▮▮▮▮▮▮" or if it could acquire the same works from LibGen without having to pay for the literary works in the ▮▮▮▮▮ catalogues. Ex. 58; *see also* Ex. 56, 59, 70, 72, 84, 92.
2022-2024 Torrent Uploads	Meta has produced no evidence that it used this data to train any Llama model.

37

META CANNOT DEFEND ITS MULTIPLE COPIES AND USES WITH ONE PURPORTEDLY FAIR USE

As alleged in the Complaint (*e.g.*, TACC ¶¶ 98, 99), and as confirmed by Meta's documents and its Rule 30(b)(6) witness in deposition, Meta made numerous copies of Plaintiffs' works for a variety of uses.

Even if Meta were permitted to make some copies for some purposes, to carry its "fair use" burden of proof, Meta was required to establish — and, for purposes of its summary judgment motion, establish without material factual dispute — that each copy and each use was fair, including that Meta only copied and used what was "reasonably necessary" for its secondary use.

META CANNOT DEFEND ITS MULTIPLE COPIES AND USES WITH ONE PURPORTEDLY FAIR USE

Meta has not remotely fulfilled its burden.

- Meta argues that a copy of Plaintiffs' works was used to train its AI models.

- Meta does not offer any evidence that all of its copies were **reasonably necessary** for training its AI models.

- Meta does not offer any evidence that all of its copies were **even used** to train its AI models.

META CANNOT DEFEND ITS MULTIPLE COPIES AND USES WITH ONE PURPORTEDLY FAIR USE

Indeed, the record evidence is contrary to what Meta must prove. For example:

- **Meta used LibGen to determine which licensing opportunities to pursue.**

  - In April 2023, after torrenting millions of copyrighted works from LibGen, Meta cross-referenced this data against publisher catalogs to determine whether it was worth pursuing licensing deals. Ex. 75 (spreadsheet comparing publisher catalogs with data found on LibGen).

- **Meta used Anna's Archive to determine which licensing opportunities to pursue.**

  - In June 2024, Meta performed "a query in union with what we collected from Anna's archive & spidermate and **will only** go after the publishing companies with the largest delta of missing content." Ex. 97 at -198.

- **Meta torrented copies of LibGen that were never incorporated into the Llama models.**

  - Meta used its entire Fall 2022 copy of LibGen for "pure exploration to see if there is value." Ex. 32. It was never incorporated into a commercially available Llama model.

- **Meta reproduced copies of millions of copyrighted works via torrenting.**

  - Meta uploaded over 40 TB of copyrighted works to internet peers in April to June 2024 alone: another "use" that is separate from LLM training. Ex. 163, Frederiksen-Cross Second Rebuttal Report, at 28-29 n. 56 ("this data shows 134.6 TB of data downloaded from the Internet, and 40.42 TB of data uploaded to the Internet").

40

FACTOR 1 FAVORS PLAINTIFFS

- Llama is a highly commercial product.

- There **are, at minimum, triable disputes** as to:

  - Whether Meta's infringement of Plaintiffs' works/copyrighted books was necessary to achieve Meta's "new purpose"; and

  - Whether Meta used Plaintiffs' works/copyrighted books for their non-expressive elements or to imitate Plaintiffs' creative expression.

FACTOR 1: WHETHER LLAMA IS A "TRANSFORMATIVE" TECHNOLOGY IS **NOT** THE CORRECT ANALYSIS

- Instead, the question is whether Meta's use of Plaintiffs' works has a further purpose "which is a matter of degree, and the degree of difference <u>must</u> be weighed against other considerations, like commercialism.'" *Warhol Found. v. Goldsmith*, 598 U.S. 508, 525 (2023) (quoting *Campbell v. Acuff-Rose Music*, Inc., 510 U.S. 569, 579 (1994)).

  - Meta's claim that "Llama's 'commercial purpose' is of little moment," (Reply Br. at 15) contravenes *Warhol* and a robust body of caselaw that precedes it.

- While Meta claims that Llama's ability to generate "entirely new text" from the data on which it has trained is dispositive of factor one, *Warhol* squarely held the opposite: "Although **new expression may be relevant** to whether a copying use has a sufficiently distinct purpose or character, **it is not, without more, dispositive of the first factor**." *Goldsmith*, 598 U.S. at 525.

FACTOR 1: META COPIED MORE THAN WAS REASONABLY NECESSARY

- *Warhol* recognizes that "a use may be justified where it is reasonably necessary to achieve the user's new purpose." 598 U.S. 508, 511 (2023).

  - Meta's own LLM expert stated that none of Plaintiffs' works impacted Llama's abilities. Instead, his experiments "demonstrate[d] the lack of any improvement over the base models beyond changes attributable to random noise." Dkt. 496, Ungar Decl., ¶ 61.

  - By Meta's own account, then, its use of Plaintiffs' works was not "reasonably necessary" to achieve Meta's "new purpose"—Plaintiffs' works did not contribute to the alleged transformativeness of Llama.

  - In Mark Zuckerberg's words: "[W]hen push comes to shove, if they demanded that we don't use their content, then we just wouldn't use their content. **It's not like that's going to change the outcome of this stuff that much.**" Adi Robertson, *Mark Zuckerberg: creators and publishers' overestimate the value' of their work for training AI,* THE VERGE, Sep. 25, 2024, at https://www.theverge.com/2024/9/25/24254042/mark-zuckerberg-creators-value-ai-meta.

    - Crediting the words of Meta's CEO, a jury could find that Meta's infringement of Plaintiffs' works was unnecessary and unjustified.

    - In fact, Meta created a model not trained on LibGen, showing that it is possible to achieve the alleged transformativeness of Llama without mass infringement.

- In contrast, Google could not have created its search engine without copyrighted books. Books were not just helpful but **necessary** to achieve Google's new purpose in creating a searchability. tool. *Google Books*, 804 F.3d at 221.

FACTOR 1: META FAILS TO SHOW THAT ITS INFRINGEMENT OF PLAINTIFFS' WORKS WAS NECESSARY TO ACHIEVE A NEW PURPOSE

- Meta cites no evidence showing that it was infeasible to develop a "transformative" LLM trained exclusively on non-infringing data. In fact, Meta also developed "non-LibGen" models but elected not to release them. Ex. 61, at -241.

- Meta cites no evidence that LLMs trained on less than full copyrighted works perform worse than LLMs trained on entire works.

- As Judge Learned Hand explained a century ago, "no plagiarist can excuse the wrong by showing how much of his work he did not pirate." *Sheldon v. Metro-Goldwyn Pictures Corp.*, 81 F.2d 49, 56 (2d Cir. 1936).

FACTOR 1: COMMERCIALITY

Meta does not dispute that Llama is a highly commercial endeavor.

- Llama powers all of Meta's commercial products: Facebook, WhatsApp, Instagram; Meta.AI.

- Meta has poured *billions* into its development of Generative AI.

  - Meta forecasts its GenAI products will *generate between $460 billion and $1.4 trillion of total revenue* by 2035. Ex. 8 at -020.

  - *See also* Ex. 12 at -008 (2024 GenAI budget showing over $900 million of Operating Expenses with plans for 2025 Operating Expenses exceeding $1 billion).

FACTOR 1: META RECOGNIZES BOOKS AS HIGHLY VALUABLE AI TRAINING DATA

Quote	Source
"Goals . . . Get as much long form writing as possible in the next 4-6 weeks."	Ex. 15, at -214
"If data quality and volumes are so important for the performance of LLMs, the best resource we can think of are definitely books."	Ex. 18, at -729
"Books are generally high-quality text."	Ex. 19, at 730
"Hence, continuing pretraining will be effective only if we have relevant self-supervised data, like books."	Ex. 16, at -954
"LibGen is the most valuable datasets that we have so far. This dataset would probably contribute the most towards improving our Agents capabilities and their creativity."	Ex. 25, at 659
"It's interesting to see how the length of documents in the datamix impacts the long context for training."	Ex. 22, at -823
It is becoming increasingly important to research the best strategies for creating long context LLMs . . . Data sources that are naturally longer[.] Books corpus (B3G or newer variants)."	Ex. 23, at -170-71
"Anna's Archive is full of long context books that they will desperately need for llama4 to have long context and there was a substantial bump in quality from libgen for similar reasons."	Ex. 24, at -041

46

FACTOR 1: META USED COPYRIGHTED BOOKS FOR THEIR EXPRESSIVE ELEMENTS

- Works such as "fictional short stor[ies]" and "motion pictures" lie closer to the core of copyright protection such that "fair use is more difficult to establish." *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 586 (1994).

- Even an instructional cake decorating booklet "involved both informational and creative aspects" and "the manner in which plaintiff assembled her book represented a creative expression." *Marcus v. Rowley*, 695 F.2d 1171, 1176 (9th Cir. 1983).

- "Computer programs differ from books, films, and many other 'literary works' in that such programs almost always serve functional purposes." *Google LLC v. Oracle America, Inc.*, 593 U.S. 1, 21 (2021).

47

FACTOR 1: META USED COPYRIGHTED BOOKS FOR THEIR EXPRESSIVE ELEMENTS

- Meta argues Llama merely "extract[s] unprotectable information about the words [books] contain." Meta Reply Br. at 16. This is wrong.

- While *Sega* and *Sony* involved reverse engineering object code to access unprotected ideas and functions, **Meta cites no case involving the "intermediate copying" of protected copyrighted expression**.

  - Meta copied the entirety of Plaintiffs' works to get the benefit of their expression.

  - Meta could have copied less of Plaintiffs' works if Meta did not intend to copy and train on Plaintiffs' creative expression.

  - Meta listed the acquisition of "creative fiction such as poems and fiction novels" as a top priority because it helped Llama with "creativity." Ex. 110 at -279.

- Meta does not dispute that Llama can mimic the style of writers; instead, it argues that "style is not protectable by copyright." Meta Reply Br. 14.

  - But Llama's ability to mimic the style of writers (including the named Plaintiffs') demonstrates that Meta has succeeded in appropriating the creativity of copyrighted works. *See* Exs. 112-114.

48

FACTOR 1: META'S "BAD FAITH" CONDUCT WARRANTS NO SKEPTICISM

- In evaluating "bad faith" under Factor One, the Ninth Circuit applies "the general rule that a party claiming fair use must act in a manner generally compatible with principles of good faith and fair dealing." *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1164 n.8 (9th Cir. 2007) (citing *Harper & Row*, 471 U.S. at 562–63).

- "Until and unless the Supreme Court rules that the defendants' good faith—or lack thereof—is not a relevant consideration, this Court is bound by the holdings of the Ninth Circuit" which include consideration of bad faith in the fair use analysis. *Triller Fight Club II LLC v. H3 Podcast*, 2023 WL 11877604, at *8 (C.D. Cal. Sept. 15, 2023).

  - **The "bad faith" evidence in this case extends well beyond subjective willful intent.** The record is replete with evidence of objective bad faith conduct that undermined the public benefits of copyright.

49

FACTOR 1: META'S "BAD FAITH" CONDUCT WARRANTS NO SKEPTICISM

- **Meta employees described LibGen as "a dataset we know to be pirated"** and feared "media coverage suggesting" Meta used that dataset "may undermine our negotiating position with regulators on these issues." Ex. 61 at -246.

  - Some Meta employees voiced concerns that "using pirated material should be beyond our ethical threshold." Ex. 32 at -171.

  - Other Meta employees noted that **"It's the piracy (and us knowing and being accomplices) that's the issue."** Ex. 156 at -906.

  - Still other Meta employees used virtual computers and Amazon Web Services for torrenting activities to "Avoid[] risk of tracing back the seeder . . . From FB server." Ex. 73 at -239.

- **Meta nevertheless conducted an "escalation to MZ" that resulted in approval of "LibGen for Llama 3."** Ex. 61 at -244.

- **Meta's own expert testified that she did not personally torrent from pirated databases** because it could "possibly either tarnish the reputation of the attorneys with whom I'm working or **expose me to risk**." Ex. 67, 163:3-6.

FACTOR 2: PLAINTIFFS' WORKS ARE ENTITLED TO THE HIGHEST COPYRIGHT PROTECTION

"Works that are creative in nature are 'closer to the core of intended copyright protection' than are more fact-based works." *Napster*, 239 F.3d at 1016; *see also Campbell*, 510 U.S. at 586 (characterizing a "fictional short story" and "motion pictures" as highly protected works).

Meta's claim that it only copied the "unprotected elements" of Plaintiffs' books, Meta Reply Br. at 16, is nonsensical: it made and used full and complete copies of Plaintiffs' works.

- Meta's conflation of the "transformation" analysis under Factor 1 with Factor 2 analysis should be rejected as a matter of law.

Nor does the fact that Plaintiffs' works were previously published weigh against them in Factor 2.

- *Every* P2P piracy case involves works that were published legitimately, then pirated onto illegal databases and torrented by infringing users.

- None of those infringements of "published" works were fair use.

FACTOR 3: META USED THE ENTIRETY OF PLAINTIFFS' WORKS

The broad test under Factor Three is whether Meta used **_only_** that portion of Plaintiffs' works that was **_necessary_** for its purportedly transformative use: "If the secondary user only copies **_as much as is necessary for his or her intended use, then this factor will not weigh against him or her_**." *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 821 (9th Cir. 2003)

- Meta misstates this straightforward test by cherry-picking courts' selectively describing the use of certain portions of work as "reasonable" – **_but that is not the test._** All of Meta's cases cited for a purported "reasonableness" test expressly put the word "reasonable" in the context of *necessity*. *See, e.g., Perfect 10*, 508 F.3d at 1167.

- Meta's cases are not to the contrary.

  - *Google Books*, 804 F.3d at 221 ("**_literally necessary_**" for Google to include total works in its search tool)

  - *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 613 (2d Cir. 2006) (author defendant's use of "reduced-size reproductions" of images of posters and tickets "displayed the *minimal image size and quality* **_necessary_** to ensure the reader's recognition" of them as "historical artifacts" of the concerts described in the book at issue).

  - In *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 640 (4th Cir. 2009), the nature of catching plagiarism by its nature called for the use of entire works.

52

FACTOR 3: META CITES NO RECORD EVIDENCE TO SUPPORT THE PURPORTED NECESSITY OF USING PLAINTIFFS' ENTIRE WORKS

Meta provides <u>no</u> record evidence to prove it was necessary to use the entirety of Plaintiffs' works (as opposed to smaller portions), or to use Plaintiffs' *particular works* at all.

- Meta cites only inapposite statements from Plaintiffs' briefs about the value of books to achieve certain benchmarks for LLMs.

- Meta contends it was entitled to use Plaintiffs' specific works as necessary for its model, but makes no factual showing why the rest of the world's textual data could not have served Llama's admittedly general-use purposes.

- *Compare Monge v. Maya Mags.*, 688 F.3d 1164, 1179 (9th Cir. 2012) (use of entirety of six copyrighted photos was **"far more than was necessary to corroborate its story"** that a couple had married); *with Bill Graham Archives*, 448 F.3d at 613 (use of whole poster and ticket images did not weigh against fair use because use of the entirety was necessary to show they were Grateful Dead "historical artifacts" associated with the band's events).

53

FACTOR 3: META'S OTHER ARGUMENTS ARE INSUFFICIENT

Meta incorrectly limits its analysis to training Llama alone, without any attempt to justify the appropriateness of its distribution or its *multiple reproductions* of *entire books* across *multiple uses* that are not, and cannot be, fair use.

Further, Meta's reliance on *Google Books* for Factor 3 is misguided.

- **The additional market here—the market for textual training data, as opposed to the sale of books as books alone—easily distinguishes this case from *Google Books*** on exactly the basis that court articulated (no market substitution by the books search tool).

- Meta's use of infringing copies of Plaintiffs' Books as training data obviously served as a **"competing substitute"** for Plaintiffs' books in the market for textual training. *E.g., Gonzalez*, 430 F.3d at 890 ("Music downloaded for free from the Internet is a close substitute for purchased music … downloading copyrighted songs cannot be defended as fair use.")

FACTOR 4:
ESTABLISHED AND POTENTIAL MARKETS

- **"A relevant derivative market includes 'those that creators of original works would in general develop or license others to develop.'"** *Dr. Seuss Enters., L.P. v. ComicMix LLC*, 983 F.3d 443, 460 (9th Cir. 2020) (quoting *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 592 (1994)).

- **The fourth factor "considers any impact on 'traditional, reasonable, or likely to be developed markets.'"** *Seltzer v. Green Day, Inc.*, 725 F.3d 1170, 1179 (9th Cir. 2013) (quoting *Ringgold v. Black Ent. Television, Inc.*, 126 F.3d 70, 81 (2d Cir. 1997)); *see also Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 930 (2d Cir. 1994) (same).

- **"To negate fair use [the plaintiff] need only show that if the challenged use should become widespread, it would adversely affect the *potential* market** for copyrighted work." *McGucken v. Pub Ocean Ltd.*, 42 F.4th 1149, 1163 (9th Cir. 2022) (quoting *Monge v. Maya Mags., Inc.*, 688 F.3d 1164, 1182 (9th Cir. 2012))

55

FACTOR 4: THE USE OF PIRATED SUBSTITUTES UNDERMINED AN "EXISTING" MARKET, A "REASONABLE" MARKET, AND/OR A "LIKELY TO BE DEVELOPED" MARKET.

- Here, a relevant market product is for *non-transformed book inputs.*

- **Courts universally recognize that direct substitution inflicts a cognizable harm on the licensing market for a copyright holder's works.** *McGucken v. Pub Ocean Ltd.,* 42 F.4th 1149, 1163 (9th Cir. 2022).

- The demand for AI training data has already been recognized as a potentially exploitable market for copyright holders. *Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc.,* 2025 WL 458520, at *9 (D. Del. Feb. 11, 2025) (recognizing potential market for copyrighted text as "data to train legal AIs") (Bibas, J.) (on appeal).

FACTOR 4: PLAINTIFFS ARE LEGALLY ENTITLED TO DEMAND COMPENSATION FOR USE OF THEIR WORKS IN THE AI DATA LICENSING MARKET

Meta's claim that Plaintiffs are not entitled to compensation for "non-substitutive transformative uses of their works" ignores the existence of the AI data licensing market and the evidentiary record establishing:

1. From the start, Meta recognized a market for licensing copyrighted material as AI training data.

2. Meta always planned to license within this market and reached out to the Big 5 and other publishers to discuss licensing.

3. When Meta realized that its pirated datasets were sufficient substitutes, it halted its licensing efforts.

4. It then adopted a "gap approach" whereby it would only license the "delta" of missing content—i.e., the content they were unable to pirate.

Whether or not Llama's "outputs" compete with Plaintiffs' works, this is not an "output" case: Meta ignores the AI training data market—a vibrant and ever-evolving market with historic precedent in machine learning where data has always been a commodity bought and sold.

FACTOR 4: TIMELINE OF PIRACY AND LICENSING

Time	Licensing	Pirating
2022	Four licenses with African-language publishers Discussions with at least one publisher for AI training data	Legal approval for exploring performance benefits of Libgen Torrented 2 million books from LibGen including 204 copies of Plaintiffs' works
Spring 2023	Potential licensing deals with ███████████████ Discussions of potential $200 million licensing budget with $100 million for books	Directive to use LibGen escalated to Mark Zuckerberg Torrented LibGen to determine whether Meta had reason to license copyrighted work beyond what LibGen contained Cross-referencing works in LibGen against catalogues offered by publishers to determine whether to license
Summer 2024	Establish "gap approach" to only license books not already contained in Meta's pirated datasets	Torrented Anna's Archive

58

FACTOR 4: BETWEEN MARCH AND APRIL 2023, META ENTERED INTO NEGOTIATIONS WITH NUMEROUS PUBLISHERS

Between October 2022 and April 2023, Meta was in talks with nearly every major book publisher to acquire copyrighted books as training data for its Llama models. Ex. 44 at -954-55 (strategy document discussing potential licensing deals for fiction with ████████████████████████████████).

Ahmad Al-Dahle (3/30/2023 10:59:00 PDT):
>We should try all the big guys:
>Harper Collins.
>Simon & Schuster.
>Macmillan.
>Hachette.
>Penguin Random House.

Ex. 41

██████	Alex B	N/A	Fiction	• Productive BD intro call on 3/16.
				• Next Steps
				   ○ Send partner evaluation agreement draft
				   ○ Sent partner supplier reactivation request.

██████	Alex B	N/A	Fiction	• AI BD having an intro call on 3/21.
██████	Alex B	N/A	Fiction	• AI BD having an intro call on 3/20.
██████	Alex B	Books		• Partner sent BD email on 3/15 that they were not interested in licensing their books for Gen AI at this time as they were still determining their strategy in this area.

Ex. 44

Between January 2023 and April 2023, as negotiations with publishers unfolded, Meta's anticipated spend on books licensing grew from $17 million to $200 million. *See* Ex. 48 at -195 ($17 million, then $27 million); Ex. 49 at -332 ("$50 million would prolly be reasonable"); Ex. 92 at -029, -030 (VP and Head of Meta's GenAI division states "It would be a travesty if our models suck because we didn't spend $100M on data.")

59

FACTOR 4: META HAS ALWAYS RECOGNIZED A MARKET FOR AI TRAINING DATA INCLUSIVE OF COPYRIGHTED BOOKS

- Meta boldly asserts "There is no market to license Plaintiffs' books for LLM training and none is likely to form." Reply at 21. This is not what the record evidences shows, and a jury could certainly find Meta has not met its burden of proof.

  - The record is replete with evidence confirming that Meta always contemplated a licensing market for copyrighted books and that Meta would "be very careful to use only licensed data for the actual training of the model." (Ex. 32).

  - In 2022, when Meta first downloaded and torrented LibGen and Books3, it was in talks with ███████ to license copyrighted data for AI training, and employees openly discussed pursuing licensing deals with publishers to acquire more copyrighted data. Ex. 32, 44.

60

FACTOR 4: META CROSS-REFERENCED LIBGEN WITH ██████████ CATALOGS TO DETERMINE WHETHER LICENSING WAS NECESSARY

Meta used LibGen to determine which licensing opportunities to pursue. After downloading millions of copyrighted works, it cross-referenced this data against publisher catalogs to determine whether it was necessary to pursue licensing deals.
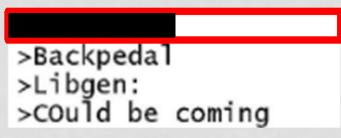
	A publisher	B title	C authors	D libgen link
1	publisher	title	authors	libgen link
2	██████	A History of Franco-German Relations in Europe	Carine Germond , Henning Türk	https://libgen.is/book/index.php?md5=B0D75615B1DAC4907113576FB8B132AA
3	██████	A History of the Royal College of General Practitioner:	John Fry, Lord Hunt of Fawley, R. J. F. H. Pinsent	https://libgen.is/search.php?req=John+H.+Hunt+%28auth.%29&column=author
4	██████	A History of UNESCO	Poul Duedahl (Professor)	https://libgen.is/search.php?req=Poul%20Duedahl%20(eds.)&column[]=author
5	██████	A Risk-Benefit Perspective on Early Customer Integra	Christoph Kausch	https://libgen.is/book/index.php?md5=879CD0BDD307EE1E36A1AADA06ED43BB
6	██████	A Primer on Spectral Theory	Bernard Aupetit	https://libgen.is/book/index.php?md5=30AED58D72082C6BE94E2ABA3C455C44
7	██████	A Resident's Guide to Psychiatric Education	Michael G. G. Thompson (Director of Education, Medical Director and Chief of	https://libgen.is/search.php?req=Michael+G.+G.+Thompson+M.D.+%28auth.%29&c
8	██████	A Time-Release History of the Opioid Epidemic	J.N. Campbell , Steven M. Rooney	https://libgen.is/book/index.php?md5=D548FEF28ACF98DF03FC295F3CFE97E8
9	██████	Back Pain in the Young Child and Adolescent	Richard M. Schwend, William L. Hennrikus	https://libgen.is/book/index.php?md5=BEF5FA23932459FC02181274613D9FAD
10	██████	Basic Concepts in Doppler Echocardiography	J. V. Chapman, A. Sgalambro	https://libgen.is/book/index.php?md5=5FF5FD8DB0C344D82CD15A25DE82C04D
11	██████	Basic Income and Sovereign Money	Geoff Crocker	
12	██████	Being and Technology	John Loscerbo	https://libgen.is/book/index.php?md5=919D29B9114BF01D8DB7A52F83DA464B
13	██████	C-C Bond Activation	Guangbin Dong	https://libgen.is/book/index.php?md5=882CA6521C228B9D69625D59A2AFB40E
14	██████	C++/CLI	Gordon Hogenson	https://libgen.is/book/index.php?md5=7C20D95125077F9B26673B812D59984C
15	██████	Campus-Management Systeme als Administrative Sy	Thorsten Spitta , Marco Carolla , Henning Brune , Thomas Grechenig , Stefan	https://libgen.is/book/index.php?md5=759663BB6382B16BBB1CBBF46BDBF6E9
16	██████	Calculations on nonlinear optical properties for large s	Feng Long Gu , Yuriko Aoki , Michael Springborg , Bernard Kirtman	https://libgen.is/book/index.php?md5=C60EE77A48005BE9436740D36C864D91
17	██████	DNA Electrophoresis	Katsuhiro Hanada	
18	██████	Dyslexia: From Theory to Intervention	Torleiv Høien , Ingvar Lundberg	https://libgen.is/book/index.php?md5=D789CB3CBDCD553EB449D09492613669

Ex. 75

The Ninth Circuit rejected the argument that using pirated data to determine what to purchase — or not purchase — is a fair use. *A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1018 (9th Cir. 2001) ("sampling remains a commercial use even if some users eventually purchase the music").

61

FACTOR 4: META CUT BACK ITS LICENSING EFFORTS AFTER DISCOVERING LIBGEN CONTAINED MOST OF THE PUBLISHERS' CATALOGS
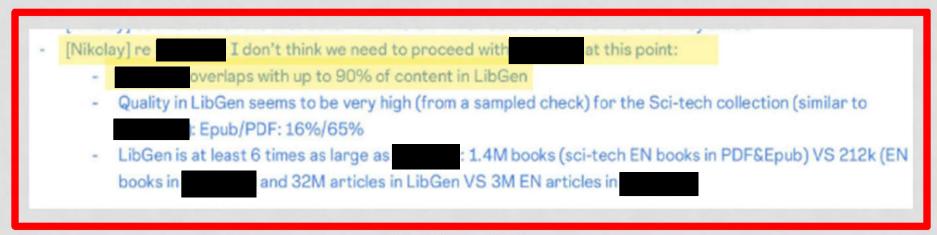
In April 2023, Meta put all licensing efforts for copyrighted books on "pause" once it realized it could obtain the same *exact* books without paying a cent.

>Backpedal
>Libgen:
>Could be coming

Ex. 107, at –772

```
8        Q.    Do you recall when you received that

9    instruction not to -- instruction not to pursue

10   any further licensing of books for --

11       A.    It was Good Friday 2023.
```

Ex. 50

- [Nikolay] re ▮▮▮▮ I don't think we need to proceed with ▮▮▮▮ at this point:
  - ▮▮▮▮ overlaps with up to 90% of content in LibGen
  - Quality in LibGen seems to be very high (from a sampled check) for the Sci-tech collection (similar to ▮▮▮▮: Epub/PDF: 16%/65%
  - LibGen is at least 6 times as large as ▮▮▮▮: 1.4M books (sci-tech EN books in PDF&Epub) VS 212k (EN books in ▮▮▮▮ and 32M articles in LibGen VS 3M EN articles in ▮▮▮▮

Ex. 56, at –018

62

FACTOR 4: META CEASED LICENSING EFFORTS AFTER DISCOVERING LIBGEN CONTAINED MOST OF THE PUBLISHERS' CATALOGS

8   A.   "▮▮▮▮▮▮▮ overlaps with up to 90% of

9   content in LibGen."

10   Q.   Okay. Is that the reason why you don't

11   think that Meta needed to proceed with ▮▮▮▮▮▮▮ at

12   that point?

13   A.   That was my understanding why I should not

14   prioritize it in my kind of scope of work to proceed

15   with analyzing or processing ▮▮▮▮▮▮▮

16   Q.   Okay. Because 90 percent of ▮▮▮▮▮▮▮ was

17   in LibGen already?

18   A.   Based on the samples that I analyzed.

Ex. 102, Bashlykov Tr., 165:6-18

63

FACTOR 4: META CONTINUES TO CROSS-REFERENCE PIRATED DATABASES AGAINST PUBLISHERS' CATALOGS

Meta's "gap approach" demonstrates market substitution.

- "For data not publicly available but considered valuable for our knowledge categories, we will initiate licensing outreach." Ex. 97, at -196.

- "As we conduct experiments/ablations, gaining more insight into the [third party] publicly available corpus, we will turn to licensing to address any shortfalls in key knowledge categories such as math, code, problem-solving, books & pdfs, and multilingual content." *Id.*

Once Meta discovered it could pirate, it lost any incentive to license. That is exactly the harm of piracy.

FACTOR 4: META CONTINUES TO CROSS-REFERENCE PIRATED DATABASES AGAINST PUBLISHERS' CATALOGS

Even after it acquired pirated books, Meta still planned to use licensing to fill in the "gaps."

Disclaimer for Publishers: for the publishing companies below we are still in the process of performing a query in union with what we collected from Anna's archive & spidermate and **will only** go after the publishing companies with the largest delta of missing content. **Redacted - Privilege**

Redacted - Privilege

2. **Licensing Data:** The strategy for this lever is detailed in this paper. We estimate that around ~10-20% of total text data corpus from Llama-4 should come from licensing.

Assumptions for Llama-4 Licensing:

[1] **Overall:** licensing alone is not a lever to meet our all data quantity goals, as no amount of reasonable licensing budget can help us get the tens of trillions of high quality tokens.

[2] **Problem solving, books, and pdfs:** categories will be heavily reliant on licensing for success given high quality content is difficult to obtain through crawling.

Ex. 97, at –192, -194, -198

Meta would not have created a detailed licensing strategy for a market that did not exist and—in its own words—was not "likely to form."

FACTOR 4: LICENSING WAS NOT COST-PROHIBITIVE

- Meta is the seventh richest company in the world, with a market cap more than **_$1.3 trillion_**. Meta's claim that licensing was cost prohibitive deserves the Court's rebuke.

- It is also contradicted by the facts:

  - Much smaller AI companies have built fairly trained models using licensed data.

  - Meta's 2024 GenAI budget was more than $900 million of with plans for 2025 Operating Expenses exceeding $1 billion. Ex. 12, at -008 (2025 GenAI Initiative).

FACTOR 4: META AND ITS AI COMPETITORS HAVE NEGOTIATED AND EXECUTED LICENSES TO TRAIN ON COPYRIGHTED MATERIAL

AI Licensing Deals for Textual Works
(As Publicly Announced and/or Reported by AAP Members)[55]

Licensee	Content Owner
Amazon	The Associated Press
Amazon	Business Insider
Amazon	Condé Nast
Amazon	Forbes
Amazon	Hearst
Amazon	Politico
Amazon	Reuters
Amazon	Time
Amazon	USA Today
Amazon	The Washington Post
Amazon	Vox
Dow Jones	The Associated Press
Dow Jones	The Wall Street Journal
Dow Jones	The Washington Post
LexisNexis	The Associated Press
Meta	Reuters
Microsoft	Axel Springer
Microsoft	Financial Times
Microsoft	HarperCollins
Microsoft	Hearst
Microsoft	Reuters
Microsoft	USA Today
Mistrial	Agence-France-Press
OpenAI	American Journalism Project
OpenAI	The Associated Press
OpenAI	The Atlantic
OpenAI	Axel Springer
OpenAI	Axios
OpenAI	Condé Nast
OpenAI	Dotdash Meredith
OpenAI	Financial Times
OpenAI	GEDI
OpenAI	Guardian Media Group

Licensee	Content Owner
OpenAI	Hearst
OpenAI	Le Monde
OpenAI	News Corp
OpenAI	Prisa Media
OpenAI	Schibsted Media Group
OpenAI	TIME
OpenAI	Vox Media
Perplexity	Adweek
Perplexity	The Independent
Perplexity	The Los Angeles Times
Potato	Wiley
ProRata.ai	Adweek
ProRata.ai	The Atlantic
ProRata.ai	Atlas Obscura
ProRata.ai	Arena Group
ProRata.ai	Axel Springer
ProRata.ai	Buzzfeed
ProRata.ai	DMG Media Group
ProRata.ai	Financial Times
ProRata.ai	Fortune
ProRata.ai	Guardian Media Group
ProRata.ai	Hello!
ProRata.ai	Mediahuis
ProRata.ai	Mumsnet
ProRata.ai	News/Media Alliance (on behalf of its members)
ProRata.ai	Prospect
ProRata.ai	Reach PLC
ProRata.ai	Sky Media Group
Confidential	Taylor & Francis
Confidential	Wiley
Confidential	Wiley
Confidential	Wiley
Confidential	Wiley

Ass'n of Am. Publishers Amicus Brief, Dkt. 535 at 12

Meta's contention that not even a *potential* market exists for books as training data is plainly incorrect—publishers and AI companies have entered into *dozens* of training data license agreements. At the very least, a jury must decide the issue.

For example, in September 2024, ███████ **licensed for training data its use of ███ books for ████████████████████ from ████████████**. Ex. 139 at -761.

This demonstrates that a market exists and that licensing books for AI training data is legally, practically, and financially feasible—especially for trillion-dollar companies like Meta and ████████.

67

FACTOR 4: PIRACY IS A KNOWN HANDICAP TO MARKET DEVELOPMENT

To the extent Meta argues no market existed or that the market was slow to develop, that impact can be largely attributed to Meta's piracy.

Similarly, it is widely recognized that the proliferation of iTunes was aided by the injunction against Napster.

See e.g., Mary Madden, *The State of Music Online: Ten Years After Napster*, Pew Research Center (June 15, 2009), at https://www.pewresearch.org/internet/2009/06/15/the-state-of-music-online-ten-years-after-napster-2/.

Meta's expert admitted he did not consider the impact of piracy on the licensing market.

```
1    BY MR. STEIN:
2         Q.   Did either of your reports look at the
3    impact of piracy on book sales?
4              MR. MORTON:  Object to form.
5              THE WITNESS:  No.
6    BY MR. STEIN:
7         Q.   Did either of your reports look at the
8    impact of the distribution aspect of torrenting on
9    book sales?
10             MR. MORTON:  Object to form.
11             THE WITNESS:  No.
```

Ex. 144, Sinkinson Tr. 257:2-11.

68

FACTOR 4: PIRATED BOOKS AND LICENSED BOOKS ARE PERFECT SUBSTITUTES

It makes no difference to the Llama model whether books were obtained via piracy or legitimate licensing agreements.

- Meta concedes that "in assessing the fourth factor, courts must determine whether the secondary use serves a 'different market function'" from the original.

- Here, the books Meta pirated for use in training Llama—without authorization or compensation—serve the *identical* market function as do licensed books.

Meta's documents show it knew that licensing a single book would effectively concede that a market existed, destroying its fair use defense:

```
Sergey Edunov (8/22/2023 11:26:06 PDT):
>the problem is that people don't realize that if we license once single book, we won't be able to lean
into fair use strategy. So we will have to drop all of the libgen and books datasets
```

Ex. 52, at -135

69